State of California—Health and Human Services Agency
California Department of Public Health

Ron Chapman, MD, MPH
Director & State Health Officer

EDMUND G. BROWN JR.
Governor

May 1, 2014

TO: Forensic Alcohol Analysis Laboratories

SUBJECT: Assigned Values and Acceptable Ranges for March 2014 Proficiency Test in Forensic Alcohol Analysis

Attached is a summary of the descriptive statistics for the March 2014 proficiency test. Included here are the target formulation values, the true values as determined by the Department's analyses, the peer-group or consensus values and the standard deviations, and a graphical summary of the distribution of participant results.

Historically, the Department has determined the acceptable limits of performance based on reported results that are within the range representing ±5% of the 99% confidence interval of the peer group mean where the range has been truncated to two significant figures (Table 1). This range is described as the "Tier #2 interval." The Department also calculates a "Tier #1 interval," which represents the range of reported results that are within ±5% of the 95% confidence interval of the peer group mean where the range is based on the results reported to three significant figures. Tier #1 is expected to include those laboratories demonstrating a high degree of accuracy. The second, wider tier would include those laboratories not as close to the central tendency as the first tier, but still accurate and therefore adequately competent. Again, historically, the Department has used the wider second tier to evaluate the laboratories' results.

The IUPAC International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories (Harmonized Protocol) recommends the use of *z*-scores for evaluating proficiency test data. With this technique, the proficiency test data are converted to a standard normal form. This is accomplished by dividing the error or difference in a reported result from the consensus value[1] by a standard deviation of the data. Various techniques have been proposed for determining the normalizing standard deviation. The Harmonized Protocol recommends the use of a "fitness-for purpose" based standard

[1] Since the consensus value is a center of the statistical distribution, several measures of central tendency can be used (mean, median, robust mean, mode). If the distribution is roughly symmetrical and unimodal then all the central tendency characteristics are coincidental. The Harmonized Protocol recommends the use of the robust mean to determine the consensus value.

deviation for proficiency assessment (σ_P). Here, fitness-for purpose is defined as the standard uncertainty that is most appropriate for the application of the results of the analysis. The Department has determined a value for σ_P based on the uncertainties associated with the reported results on recent tests together with the 5% accuracy and precision standard of performance requirements set forth in the regulations [cf. Title 17 §1220.1.(a)(1)]. The Department found that a reasonable value for σ_P expressed as a relative standard deviation is 2.5%. The standard deviation can also be determined based on the data obtained from a given proficiency test round. The Harmonized Protocol does not recommend this procedure noting that this can result in variations in the determinations of acceptable limits from test to test. Z-score can be thus calculated using the formula and data provided in Table 2.

$$z = \frac{\mathrm{X} - \mathrm{X}_a}{\sigma_p}$$

Where:

X - individual laboratory results for a given pool
X_a - consensus value
σ_p – fitness-for-purpose standard deviation

The primary advantage of *z*-scores is that they make all proficiency test results directly comparable regardless of concentration. A laboratory's performance can be easily interpreted from a *z*-score. Generally a score between -2 and +2 ($|z| \leq 2$) is considered satisfactory or acceptable. A *z*-score outside the range -3 to +3, inclusive ($|z| \geq 3$) is considered unsatisfactory or unacceptable and the laboratory must take corrective actions. *Z*-scores between -3 and -2 or +2 and +3 ($2 < |z| < 3$) are considered questionable and these two ranges should be used as warning limits. Scores within the warning limit ranges in two or more consecutive test events could be considered unacceptable.

Sincerely,

Clay Larson, Chief
Abused Substances Analysis Section
Food and Drug Laboratory Branch

Statistical Data for March 2014 Proficiency Test in Forensic Alcohol Analysis

Table 1 CDPH Tier#1 and Tier #2 Acceptable Ranges

Pool	Peer Group Mean	Tier #1	Tier #2
#1	0.143	0.134 – 0.152	0.13 – 0.15
#2	0.253	0.237 – 0.269	0.23 – 0.27

Table 2 Suimmary of Test Pool Data

Parameter	POOL 1 (02244)	POOL 2 (03034)
Pre-distribution Data	Target Value 0.14% True Value[2] 0.140 Standard Deviation[2] 0.0003	Target Value 0.25% True Value[2] 0.250 Standard Deviation[2] 0.0013
Descriptive statistics	Mean 0.142 Adjusted Mean[3] 0.143 Standard Error[4] 0.0005 Median 0.143 Standard Deviation 0.0033 Minimum 0.135 Maximum 0.149 Count 49	Mean 0.253 Adjusted Mean[4] 0.253 Standard Error[5] 0.0007 Median 0.253 Standard Deviation 0.0046 Minimum 0.240 Maximum 0.261 Count 47
Histogram	Figure 1	Figure 2
Normal distribution?[5]	NO	YES
Robust mean, X*	0.143	0.253
Robust standard deviation, σ_{rob}	0.0031	0.0044
Fitness-for-purpose standard deviation, σ_p	0.0036	0.0063
Consensus value (X_a) determined as Mode ($\mu_{1/2}$)	0.1434	0.2536
Uncertainty of the consensus value, X_a , S.E.[6]	0.0005	0.0008
$X_a \pm$ S.E.	0.143 ± 0.0005	0.254 ± 0.0008
z-score	$z = \frac{X - X_a}{\sigma_p}$	$z = \frac{X - X_a}{\sigma_p}$

[2] Based on CDPH's Headspace Gas Chromatographic Method
[3] Mean determined from participant data after the removal of outlier(s)
[4] Standard Error of the Mean
[5] Shapiro-Wilk test used
[6] Determined as Standard Error of Mode using bootstrap simulation technique with bandwidth of $0.75*\sigma_p$

Figure 1

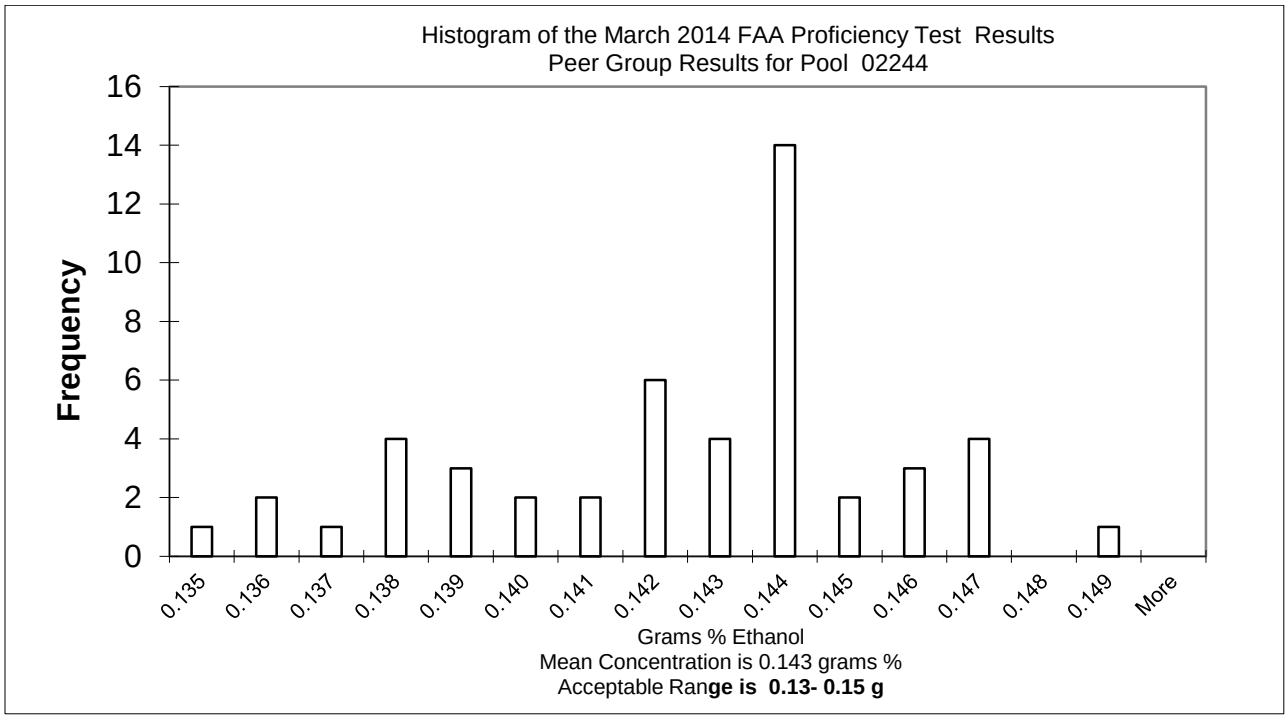

Figure 2